

```
In [2]: import pandas as pd
from sklearn import preprocessing
df=pd.read_csv("/home/student/Desktop/C0TC50/Iris.csv")
df
```

```
Out[2]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

```
In [2]: df.mean()
```

/tmp/ipykernel_5352/3698961737.py:1: FutureWarning: The default value of numeric_only in DataFrame.mean is deprecated. In a future version, it will default to False. In addition, specifying 'numeric_only=None' is deprecated. Select only valid columns or specify the value of numeric_only to silence this warning.

```
df.mean()
Out[2]: Id          75.500000
SepalLengthCm    5.843333
SepalWidthCm     3.054000
PetalLengthCm    3.758667
PetalWidthCm     1.198667
dtype: float64
```

```
In [3]: df.loc[:, "SepalLengthCm"].mean()
```

```
Out[3]: 5.8433333333333334
```

```
In [6]: df.mean(axis=1)[0:4]
```

/tmp/ipykernel_5352/1148177455.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

```
df.mean(axis=1)[0:4]
Out[6]: 0      2.24
1      2.30
2      2.48
3      2.68
dtype: float64
```

```
In [7]: df.median()
```

```
/tmp/ipykernel_5352/530051474.py:1: FutureWarning: The default value of
numeric_only in DataFrame.median is deprecated. In a future version, it
will default to False. In addition, specifying 'numeric_only=None' is d
eprecated. Select only valid columns or specify the value of numeric_on
ly to silence this warning.
```

```
df.median()
Out[7]: Id          75.50
SepalLengthCm    5.80
SepalWidthCm     3.00
PetalLengthCm    4.35
PetalWidthCm     1.30
dtype: float64
```

```
In [8]: df.loc[:, "SepalLengthCm"].median()
```

```
Out[8]: 5.8
```

```
In [9]: df.median(axis=1)[0:4]
```

```
/tmp/ipykernel_5352/381455229.py:1: FutureWarning: Dropping of nuisance
columns in DataFrame reductions (with 'numeric_only=None') is deprecate
d; in a future version this will raise TypeError. Select only valid co
lumns before calling the reduction.
```

```
df.median(axis=1)[0:4]
Out[9]: 0    1.4
1    2.0
2    3.0
3    3.1
dtype: float64
```

```
In [10]: df.mode()
```

```
Out[10]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.0	3.0	1.5	0.2	Iris-setosa
1	2	NaN	NaN	NaN	NaN	Iris-versicolor
2	3	NaN	NaN	NaN	NaN	Iris-virginica
3	4	NaN	NaN	NaN	NaN	NaN
4	5	NaN	NaN	NaN	NaN	NaN
...
145	146	NaN	NaN	NaN	NaN	NaN
146	147	NaN	NaN	NaN	NaN	NaN
147	148	NaN	NaN	NaN	NaN	NaN
148	149	NaN	NaN	NaN	NaN	NaN
149	150	NaN	NaN	NaN	NaN	NaN

150 rows × 6 columns

```
Out[11]: 0      5.0
          Name: SepalLengthCm, dtype: float64
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
'<' not supported between instances of 'str' and 'int'
df.mode(axis=1)[0:4]
/tmp/ipykernel_17236/657814072.py:1: UserWarning: Unable to sort modes:
'<' not supported between instances of 'str' and 'int'
df.mode(axis=1)[0:4]
/tmp/ipykernel_17236/657814072.py:1: UserWarning: Unable to sort modes:
'<' not supported between instances of 'str' and 'int'
df.mode(axis=1)[0:4]
/tmp/ipykernel_17236/657814072.py:1: UserWarning: Unable to sort modes:
'<' not supported between instances of 'str' and 'int'
df.mode(axis=1)[0:4]
/tmp/ipykernel_17236/657814072.py:1: UserWarning: Unable to sort modes:
'<' not supported between instances of 'str' and 'int'
df.mode(axis=1)[0:4]
```

```
Out[4]:
```

	0	1	2	3	4	5	
0	1	5.1	3.5	1.4	0.2		Iris-setosa
1	2	4.9	3.0	1.4	0.2		Iris-setosa
2	3	4.7	3.2	1.3	0.2		Iris-setosa
3	4	4.6	3.1	1.5	0.2		Iris-setosa

```
In [15... df.min()
```

```
Out[15]:
```

Id	1
SepalLengthCm	4.3
SepalWidthCm	2.0
PetalLengthCm	1.0
PetalWidthCm	0.1
Species	Iris-setosa

dtype: object

```
In [16... df.loc[:, "SepalLengthCm"].min(skipna = False)
```

```
Out[16]: 4.3
```

```
In [17... df.max()
```

```
Out[17]:
```

Id	150
SepalLengthCm	7.9
SepalWidthCm	4.4
PetalLengthCm	6.9
PetalWidthCm	2.5
Species	Iris-virginica

dtype: object

```
In [18... df.loc[:, "SepalLengthCm"].max(skipna = False)
```

```
Out[18]: 7.9
```

```
In [19... df.std()
```

```
/tmp/ipykernel_5352/3390915376.py:1: FutureWarning: The default value of
numeric_only in DataFrame.std is deprecated. In a future version, it
will default to False. In addition, specifying 'numeric_only=None' is d
eprecated. Select only valid columns or specify the value of numeric_on
ly to silence this warning.
```

```
df.std()
Out[19]:
```

Id	43.445368
SepalLengthCm	0.828066

```
SepalWidthCm      0.433594
PetalLengthCm     1.764420
PetalWidthCm      0.763161
dtype: float64
```

```
In [21]: df.loc[:, 'SepalLengthCm'].std()
```

```
Out[21]: 0.8280661279778629
```

```
In [22]: df.std(axis=1)[0:4]
```

```
/tmp/ipykernel_5352/3966588610.py:1: FutureWarning: Dropping of nuisance
columns in DataFrame reductions (with 'numeric_only=None') is depreca
ted; in a future version this will raise TypeError.  Select only valid
columns before calling the reduction.
```

```
Out[22]: df.std(axis=1)[0:4]
0      2.010721
1      1.772005
2      1.754138
3      1.813009
dtype: float64
```

```
In [23]: df.groupby(['Id'])['SepalLengthCm'].mean()
```

```
Out[23]: Id
1      5.1
2      4.9
3      4.7
4      4.6
5      5.0
...
146     6.7
147     6.3
148     6.5
149     6.2
150     5.9
Name: SepalLengthCm, Length: 150, dtype: float64
```

```
In [29]: df_u=df.rename(columns= {'SepalLengthCm$':'SepalLen'},inplace=False)
```

```
In [32]: enc = preprocessing.OneHotEncoder()
enc_df = pd.DataFrame(enc.fit_transform(df[['Species']]).toarray())
enc_df
```

```
Out[32]:
```

	0	1	2
0	1.0	0.0	0.0
1	1.0	0.0	0.0
2	1.0	0.0	0.0
3	1.0	0.0	0.0
4	1.0	0.0	0.0
...
145	0.0	0.0	1.0
146	0.0	0.0	1.0
147	0.0	0.0	1.0
148	0.0	0.0	1.0

```
      0    1    2
```

```
149  0.0  0.0  1.0
```

150 rows × 3 columns

```
In [33... df_encode=df_u.join(enc_df)
df_encode
```

```
Out[33]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	0	1
0	1	5.1	3.5	1.4	0.2	Iris-setosa	1.0	0.0
1	2	4.9	3.0	1.4	0.2	Iris-setosa	1.0	0.0
2	3	4.7	3.2	1.3	0.2	Iris-setosa	1.0	0.0
3	4	4.6	3.1	1.5	0.2	Iris-setosa	1.0	0.0
4	5	5.0	3.6	1.4	0.2	Iris-setosa	1.0	0.0
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica	0.0	0.0
146	147	6.3	2.5	5.0	1.9	Iris-virginica	0.0	0.0
147	148	6.5	3.0	5.2	2.0	Iris-virginica	0.0	0.0
148	149	6.2	3.4	5.4	2.3	Iris-virginica	0.0	0.0
149	150	5.9	3.0	5.1	1.8	Iris-virginica	0.0	0.0

150 rows × 9 columns

```
In [35... col_names = ['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm']
```

```
In [43... iris = pd.read_csv("/home/student/Desktop/C0TC50/Iris.csv")
```

```
In [44... irisSet = (iris['Species'] == 'Iris-setosa')
```

```
In [47... iris[irisSet] == "Iris-setosa"
print('Iris-setosa')
print(iris[irisSet].describe())
```

Iris-setosa

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	50.00000	50.00000	50.000000	50.000000	50.00000
mean	25.50000	5.00600	3.418000	1.464000	0.24400

std	14.57738	0.35249	0.381024	0.173511	0.107
21					
min	1.00000	4.30000	2.300000	1.000000	0.100
00					
25%	13.25000	4.80000	3.125000	1.400000	0.200
00					
50%	25.50000	5.00000	3.400000	1.500000	0.200
00					
75%	37.75000	5.20000	3.675000	1.575000	0.300
00					
max	50.00000	5.80000	4.400000	1.900000	0.600
00					

```
In [54... irisVer = (iris['Species']== 'Iris-versicolor')
print('Iris-versicolor')
print(iris[irisVer].describe())
```

Iris-versicolor					
	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
hCm					
count	50.00000	50.000000	50.000000	50.000000	50.000
000					
mean	75.50000	5.936000	2.770000	4.260000	1.326
000					
std	14.57738	0.516171	0.313798	0.469911	0.197
753					
min	51.00000	4.900000	2.000000	3.000000	1.000
000					
25%	63.25000	5.600000	2.525000	4.000000	1.200
000					
50%	75.50000	5.900000	2.800000	4.350000	1.300
000					
75%	87.75000	6.300000	3.000000	4.600000	1.500
000					
max	100.00000	7.000000	3.400000	5.100000	1.800
000					

```
In [56... irisVir = (iris['Species']== 'Iris-virginica')
print('Iris-virginica')
print(iris[irisVir].describe())
```

Iris-virginica					
	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
hCm					
count	50.00000	50.000000	50.000000	50.000000	50.000
000					
mean	125.50000	6.588000	2.974000	5.552000	2.02
600					
std	14.57738	0.63588	0.322497	0.551895	0.27
465					
min	101.00000	4.900000	2.200000	4.500000	1.40
000					
25%	113.25000	6.225000	2.800000	5.100000	1.80
000					
50%	125.50000	6.500000	3.000000	5.550000	2.00
000					
75%	137.75000	6.900000	3.175000	5.875000	2.30
000					
max	150.00000	7.900000	3.800000	6.900000	2.50
000					

In []: